Use of Artificial Intelligence for the Preoperative Diagnosis of Pulmonary Lesions

Fred H. Edwards, MD, Paul S. Schaefer, MD, Amram J. Cohen, MD, Ronald F. Bellamy, MD, LeNardo Thompson, MD, Geoffrey M. Graeber, MD, and Michael J. Barry, MD

Department of Cardiothoracic Surgery, Walter Reed Army Medical Center, Washington, DC, the Walter Reed Institute of Research, Washington, DC, and the F. Edward Hebert School of Medicine, Uniformed Services University of Health Sciences, Bethesda, Maryland

The relatively new field of artificial intelligence has spawned a variety of techniques associated with computer-assisted diagnosis. These techniques have been applied to the diagnosis of pulmonary lesions, but previous reports have focused on medical rather than surgical populations and the results have been evaluated using only retrospective patient surveys. We used a Bayesian algorithm to develop a diagnostic computer model for prospectively evaluating patients undergoing thoracotomy for suspected pulmonary malignancy. Patients who had a preoperative diagnosis were not included. Preoperative clinical and radiographic parameters for 100 consecutive patients were prospectively entered into the diagnostic model, which then categorized the lesion as benign or malignant. The computer predictions agreed with the final histological diagnosis in 95 of the 100 patients. The sensitivity was 96% and the specificity was 89% for this prospective series. These results indicate that the computer-assisted diagnosis of pulmonary lesions may have a role in this clinical setting.

(Ann Thorac Surg 1989;48:556–9)

There are well-accepted clinical and radiographic parameters associated with malignancy, but the diagnosis of new pulmonary lesions is nevertheless quite challenging. Because of this difficulty in making an accurate preoperative diagnosis of such lesions, patients may undergo unnecessary invasive procedures and may encounter some delay before receiving treatment.

Recently, artificial intelligence techniques have been applied to this problem [1–4]. Most approaches deal with computer-assisted diagnosis based on expert systems or Bayesian theory, and several studies have reported accurate results [1–4]. The goal of these studies is to provide an accurate diagnostic adjunct that can be integrated with more conventional information to arrive at a clinical diagnosis.

We have used a Bayesian model to predict the diagnosis of newly discovered pulmonary lesions. Our initial experience involved the prospective evaluation of patients undergoing thoracotomy for suspected bronchogenic cancer. The model produced a 96% diagnostic accuracy, but more than one third of these patients had a histological diagnosis before thoracotomy, indicating that the model held promise for practical clinical application. The true test, however, must address the diagnosis of patients in whom the preoperative diagnosis is not known. The purpose of the present study was to use computer-assisted diagnosis to categorize pulmonary lesions as being benign or malignant in a population of patients in whom the preoperative diagnosis had not been established.

Theory

We previously described a step-by-step approach [5] that serves as a technical guide for physicians and other individuals who wish to develop Bayesian studies. The more intricate details of Bayesian theory are covered in that reference, but a general overview of this approach may be helpful to introduce our present work.

Bayesian algorithms use clinical observations of previously evaluated patients to predict the diagnosis of new patients. These clinical observations usually take the form of risk factors that are selected for their ability to discriminate between diagnostic categories. The frequency with which the risk factors are found in each diagnostic category makes up a conditional probability matrix (CPM) that is incorporated into a computerized Bayesian algorithm.

The conditional probabilities are derived from retrospective patient surveys, published data, or physician estimates [1, 2, 5, 6]. The technique used to generate the CPM determines the clinical data base on which diagnostic predictions are made. For that reason, a retrospective patient survey from one's own institution ensures that the model is tailored to reflect the unique experience of that institution.

After the model has been developed, a new patient can be evaluated by tabulating the presence or absence of each risk factor for that patient. The Bayesian algorithm then uses the clinical experience embodied in the CPM to

Accepted for publication June 26, 1989.

Address reprint requests to Dr Edwards, Department of Thoracic Surgery, Walter Reed Army Medical Center, Washington, DC 20307-5001.

calculate the probability that the patient will fall into a given diagnostic category.

Material and Methods

In this study, two diagnostic categories were considered: "benign" and "malignant." The risk factors (Table 1) were restricted to clinical and radiographic parameters that are readily available as part of the routine preoperative evaluation for patients suspected of having pulmonary malignancy. We selected the factors that we considered important in discriminating between benign and malignant lesions [1, 4, 7]. The associated conditional probabilities were derived from a combination of physician estimates and a retrospective review of our clinical experience. This information was used to develop a computerized Bayesian algorithm that applied the CPM to predict the diagnosis of new patients.

The patient population for entry into this model was drawn from the recent operative experience at our institution. From January 1986 to January 1988, 165 consecutive patients underwent thoracotomy for suspected pulmonary malignancy. All patients underwent our usual preoperative diagnostic evaluation. Chest roentgenograms were obtained for each patient, and most had computed tomography of the chest.

Each patient underwent fiberoptic bronchoscopy and, if an endobronchial lesion was identified, a biopsy specimen was taken. Otherwise, brushings and washings of appropriate areas were performed. Generally, an attempt was made to perform a transthoracic needle biopsy if the lesion was in a peripheral location. Patients with lesions believed to be accessible by transbronchial biopsy also underwent that procedure. If computed tomography identified mediastinal lymph nodes larger than 1.5 cm in diameter, a staging procedure with cervical mediastinoscopy or anterior mediastinotomy was performed.

Using this approach, we confirmed a preoperative diagnosis of cancer in 65 patients. A preoperative diagnosis could not be established in the remaining 100 patients, who constituted the study population. A posterolateral thoracotomy was performed on each patient, and the pulmonary lesion was completely excised and submitted for histological examination.

The presence or absence of the risk factors (Table 1) was entered into the model for each patient, and the Bayesian algorithm then calculated the probability that the lesion was benign or malignant. The "diagnostic prediction" for a given patient was the alternative (benign or malignant) calculated to have the higher probability. To test the validity of the diagnostic model, this calculated result was compared with the final histological diagnosis obtained from the excised specimen.

Results

Table 1 shows a clinical profile of the patient population. Of the 100 patients undergoing thoracotomy, 82 had malignant lesions and the remaining 18 had benign lesions. The model correctly categorized the lesion as benign or malignant in 95 of the 100 cases, yielding a 95%

Table 1. Preoperative Patient Characteristics

No. of Patients	Total (n = 100)	Cancer[a] (n = 82)	Benign[a] (n = 18)
Age < 45 yr	16	10 (12)	6 (33)
Age ≥ 45 yr	84	72 (88)	12 (67)
Male	71	56 (68)	15 (83)
Female	29	26 (32)	3 (17)
Smoking history	81	69 (84)	12 (67)
Weight loss	13	13 (16)	0 (0)
Hemoptysis	11	10 (12)	1 (6)
New cough	33	31 (38)	2 (11)
Chest pain	7	6 (7)	1 (6)
HPO symptoms	3	3 (4)	0 (0)
Bloody effusion	0	0 (0)	0 (0)
Previous cancer	25	24 (29)	1 (6)
Lesion < 3 cm	47	32 (39)	15 (83)
Lesion 3.1–5 cm	41	39 (48)	2 (11)
Lesion > 6 cm	12	11 (13)	1 (6)
Cavitation (thick wall)	8	7 (9)	1 (6)
Cavitation (thin wall)	1	0 (0)	1 (6)
Smooth margins	23	8 (10)	15 (83)
Slightly irregular margins	42	41 (50)	1 (6)
Very irregular margins	29	28 (34)	1 (6)
Lobulation	10	9 (11)	1 (6)
Spiculation	15	15 (18)	0 (0)
Homogeneous	70	57 (70)	13 (72)
Inhomogeneous	26	23 (28)	3 (17)
Tracheal deviation	3	3 (4)	0 (0)
No calcification	95	78 (95)	17 (94)
Eccentric calcification	0	0 (0)	0 (0)
Central/popcorn calcification	0	0 (0)	0 (0)
Atelectasis	20	19 (23)	1 (6)
Chest wall abutment	16	14 (17)	2 (11)
Chest wall invasion	2	2 (2)	0 (0)
Mediastinal abutment	14	13 (16)	1 (6)
N1 enlargement	9	9 (11)	0 (0)
N2 enlargement	11	11 (13)	0 (0)
No size increase in 6 mo	4	1 (1)	3 (17)
Size increase in 6 mo	49	44 (54)	5 (28)
Effusion on chest film	3	3 (4)	0 (0)
Middle lobe lesion	5	2 (2)	3 (17)
Bilateral lesions	9	8 (10)	1 (6)
Multiple unilateral lesions	5	4 (5)	1 (6)
No lesions on FOB	69	52 (63)	17 (94)
Endobronchial lesion seen	4	4 (5)	0 (0)
Extrinsic compression on FOB	2	2 (2)	0 (0)
Blunted carina on FOB	0	0 (0)	0 (0)

[a] Numbers in parentheses are percentages.

FOB = fiberoptic bronchoscopy; HPO = hypertrophic pulmonary osteoarthropathy.

Table 2. Confidence Limits of Selected Indexes

Index	90% Confidence Level (%)	70% Confidence Level (%)
Accuracy (95%)	89–98	92–97
Sensitivity (96%)	90–99	93–98
Specificity (89%)	68–98	76–96
Predictive value of		
Positive test (97%)	92–99	94–99
Negative test (84%)	63–95	71–93

accuracy. Sixteen of the 18 benign lesions and 79 of the 82 malignant lesions were correctly assigned, thereby producing two false-positive and three false-negative results. This yielded a sensitivity of 96% and a specificity of 89% with confidence limits as shown in Table 2.

The predictive value of a positive test is the probability that a patient with a positive test does in fact have the disease in question [8]. In the present context, a positive test refers to the Bayesian prediction of cancer and the disease in question is pulmonary malignancy. The predictive value of a positive test was 97% in this series. The predictive value of a negative test [8], ie, the probability that a patient with a benign test result has a benign pulmonary lesion, was 84% for this model.

The likelihood ratio [8] is a useful entity that provides an intuitive measure of test accuracy. In this study, the likelihood ratio for a positive test was 8.7, which indicates that a person with cancer is 8.7 times more likely to test positive than a person with benign disease. Conversely, the likelihood ratio for a negative test is 0.04, indicating that a person with cancer is 0.04 times as likely to test negative as compared with a person with a benign process.

The results discussed apply to a discrete "yes or no" application of the diagnostic algorithm. Actually, the model calculates the probability of each diagnostic alternative. We arbitrarily selected the alternative with the higher probability as the computer diagnosis, but examination of a breakdown of the actual calculated probabilities for these patients is useful.

Table 3 shows the results for calculating the probability of a benign lesion. There was reasonable agreement between the observed and predicted results, but the most salient feature of Table 3 concerns the patients predicted to have a very low probability of benign disease. The model predicted that the probability of having a benign lesion was less than 5% for 57 patients. In fact, none of these patients had benign disease. Conversely, all 57 patients with greater than 95% probability of having a malignant lesion were found to have cancer at the time of operation.

Comment

Several studies have reported the use of computerized mathematical algorithms to assist in the diagnosis of pulmonary lesions. As the field of computer-assisted diagnosis has evolved, it has become apparent that Bayesian theory is most suited to this task; in fact, all successful applications have used this technique [1, 4–6, 9]. Earlier reports examined the entire spectrum of patients with pulmonary lesions, but we focused on a more select and more challenging subgroup. By restricting our analysis to patients undergoing operation without a preoperative diagnosis, we directly addressed the most compelling practical problem in this clinical context.

Surprisingly, none of the previous studies used a prospective analysis to validate the diagnostic model. Instead, the test group was the same population of patients that was used to generate the model [1, 2, 9]. Clearly, the goal of diagnostic mathematical algorithms should be to predict the diagnosis of new patients rather than to model previously evaluated patients. In the present study, the Bayesian algorithm was developed before our prospective test group was examined so that the evaluation of these 100 prospective patients was completely independent of the patient population used to derive the model.

The model we present has proved to be an accurate diagnostic tool that other researchers may find helpful. However, a program developed in one institution may not be directly applicable to other institutions. The way in which the CPM is developed will effectively tailor the model to reflect the clinical experience of that institution. This is both an advantage and a disadvantage to potential users. The disadvantage is that a program developed in one practice cannot usually be used in another practice because of the differences in patient population and differences in the approach to preoperative evaluation. For example, a group that routinely performs mediastinoscopy on all patients will not have the same mix of patients undergoing thoracotomy as we do, nor will those groups that only rarely use mediastinoscopy. The advantage is that one can create a diagnostic model that conforms closely to the unique approach that is used in a given institution. Most surgeons are aware of the difficulties inherent in extrapolating information from reported series for use in their own practice. Problems of this kind can be obviated by the ability to tailor the model to reflect the specific population of any given hospital.

Table 3. Predicted Versus Observed Results for Benign Lesions

Predicted Probability of Benign Lesions (%)	Observed Frequency of Benign Lesions[a]
<5	0% (0/57)
5–25	7% (1/15)
25–50	11% (1/9)
50–75	67% (4/6)
>75	92% (12/13)

[a] In parentheses, the denominator is the number of patients with the predicted probability of benign disease that lies within the range shown in the first column. The numerator is the number of patients who actually had a benign lesion (eg, 13 patients were predicted to have >75% probability of benign disease and 12 of those patients did have a benign lesion).

The risk factors for malignancy (Table 1) were selected from our own clinical observations and from reports in the literature [1, 2, 4, 7]. We do not contend that these parameters are the only ones of importance, nor do we contend that all of them are necessary in such an analysis. One of the advantages of Bayesian theory is that one does not pay a penalty for selecting a risk factor that has little impact on the diagnosis. If a parameter does not significantly discriminate between diagnostic categories, the derived conditional probabilities for that factor will be approximately equal for each diagnosis, so that the factor simply has little mathematical consequence in the final calculations [5, 6, 9]. Because of this, we have been liberal in selecting our preoperative clinical and radiographic parameters, and we encourage other researchers to adopt a similar philosophy. The choice of these factors is discussed in some detail in other reports [3–6], which may be useful to potential investigators.

We have shown that computer-assisted diagnosis can provide accurate results in the preoperative assessment of pulmonary lesions, but the question of how to use this information is still unsettled. Certainly any test with more than 90% accuracy should be welcome in this clinical setting, particularly if the test is not invasive and has no morbidity or cost.

The utility of our approach is supported by the fact that the sensitivities, specificities, and predictive accuracy of the results were high both in the present study and in our previous work [4], even though the test populations were different and the sample populations were only partially overlapping.

This type of test does not dictate therapy [3–5, 9], but rather serves as an adjunct that should be evaluated with other preoperative tests to arrive at a clinical diagnosis. We have not used the test to influence our preoperative management, but such consideration may be warranted. If the model predicts that the probability of cancer is greater than 95% and our clinical judgment is consistent with a malignant lesion, perhaps it would be appropriate to perform thoracotomy without additional procedures that might otherwise have been done. Under these circumstances, the test provides an objective, statistically rigorous basis for our approach. Even this application may appear radical to some investigators, but we often use other tests, such as transthoracic needle biopsy, that are more invasive, more morbid, more costly, and less accurate.

References

1. Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules. Am Rev Respir Dis 1986;134:449–52.
2. Rotte KH, Meiski W. Results of computer-aided diagnosis of peripheral bronchial carcinoma. Radiology 1977;125:583–6.
3. Rogers W, Ryak B, Moeller G. Computer-aided diagnosis: literature review. Int J Biomed Comp 1979;19:267–70.
4. Edwards FH, Schaefer P, Callahan S, et al. Bayesian statistical theory in the preoperative diagnosis of pulmonary lesions. Chest 1987;92:888–91.
5. Edwards FH, Graeber GM. The theorem of Bayes as a clinical research tool. Surg Gynecol Obstet 1987;165:127–9.
6. Raeside DE. Bayesian statistics—a guided tour. Med Phys 1976;3:1–11.
7. Godwin J. The solitary pulmonary nodule. Radiol Clin North Am 1983;21:709–21.
8. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. Biostatistics in clinical medicine. New York: Macmillan, 1987:7–32.
9. Wardle A, Wardle L. Computer-aided diagnosis—a review of research. Methods Inform Med 1978;17:15–28.